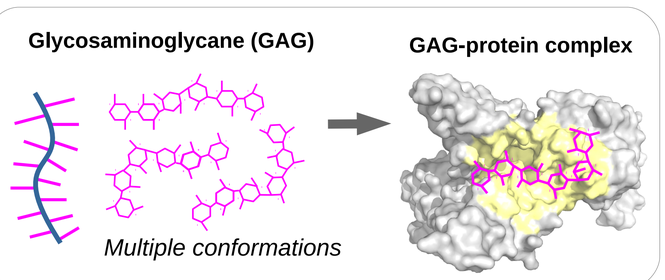
Fragment-based modeling of protein-GAG complexes

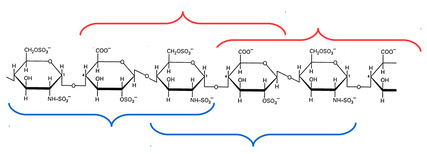
Isaure Chauvot de Beauchêne [1] Sergey Samsonov [2] Martin Zacharias [3]

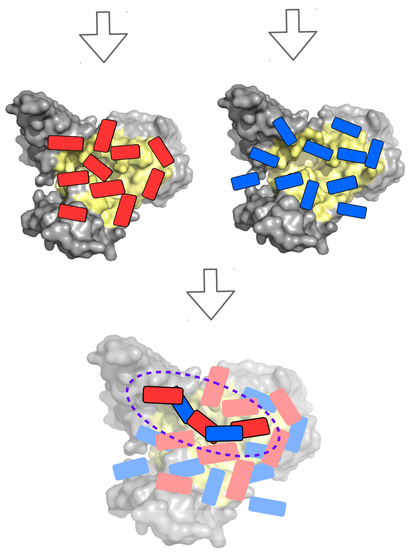
[1] CNRS, LORIA, Nancy [2] University of Gdańsk [3] Technic Universiy Munich

INTRODUCTION

Glycosaminoglycans (GAGs) are **linear anionic periodic poly-saccharides**. They bind to their protein targets in the extracellular matrix, and so participate in many cell-signaling processes [1]. As such, they are very promising targets for the design of novel functional **biomaterials** for **regenerative medicine** [2]







Experimental methods (X-ray, NMR) to solve the structure of GAG-protein complexes are impaired by GAG flexibility and high negative charge density that makes it difficult to obtain pure homogeneous samples. Therefore, it often requires computational docking methods. But GAGs are also **too flexible for (semi-)rigid docking, and too large for fully flexible docking.**

Fragment-based docking permits to deal with the high number of degrees of freedom of the ligand. However, most current methods are limited to small ligands and well delineated binding sites. We have recently extended such approach to flexible linear polymers docked on the whole protein surface [3].

By rigid docking of fragment libraries and exhaustive assembly of the compatible poses, we could model protein-bound ssRNA, another highly flexible anionic linear polymer [4]. We present here a new variation on the thema, using **flexible docking** of fragments by AutoDock, on a coarsly **known binding site**. We sucessfully applied it on protein – GAG complexes.

RESULTS

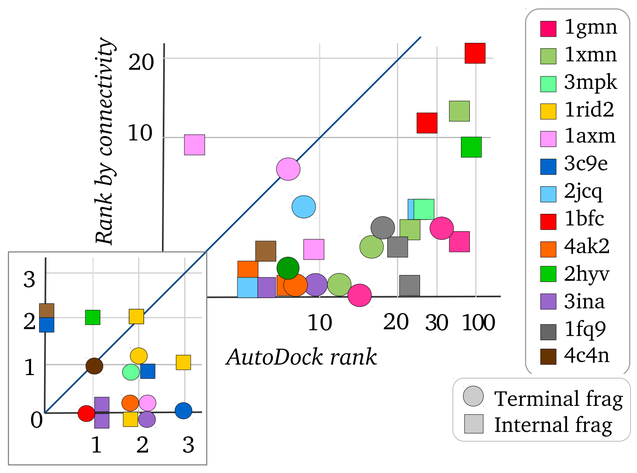
Protein-GAG benchmark

dp = degree of polymerisation

Pdb id	protein	GAG	dp
2axm	FGF1	heparin	dp6
1bfc	FGF2	heparin	dp6
1fq9	FGF2-FGFR1	heparin	dp6
1gmn	NK1 (HGF)	heparin	dp5
1rid	VCP	heparin	dp7
1xmn	Thrombin	heparin	dp6
2hyv	Annexin 2A	heparin	dp5
2jcq	CD44	hyaluronan	dp6
3c9e	Cathepsin K	chondroitin sulfate	dp6
3ina	Heparinase I	heparin	dp7
3mkp	VFT2	heparin	dp5
4ak2	Suse-like	heparin	dp6
4c4n	Hedgehog	heparin	dp6

Chain-assembly is especially suited to select terminal fragments, which AD rank worse than poses located at the binding site of a central fragment.

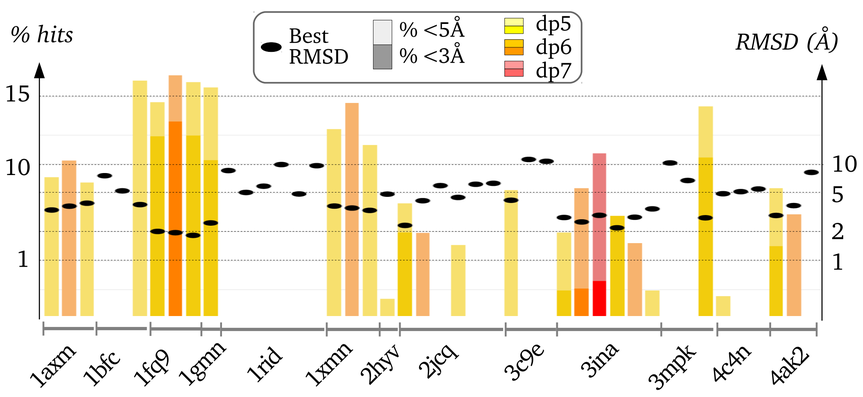
Hits enrichment by fragment assembly



Filtering the poses by fragment assembly propensity (= counting the possible chains they belong to) enriches the pool in native poses (<5Å) **more effectively than AD scoring**.

However, this filtering did not permit to precise the binding site prediction by counting the number of contacts per amino-acid.

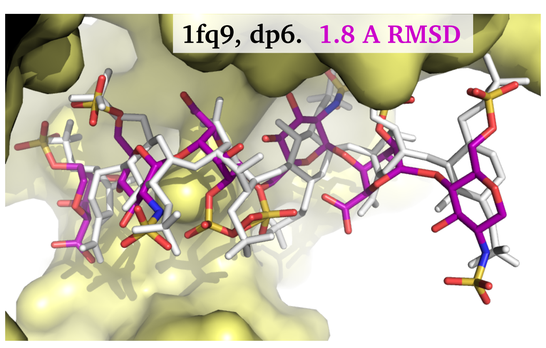
Chain assembly for GAGs with different length (*dp*)

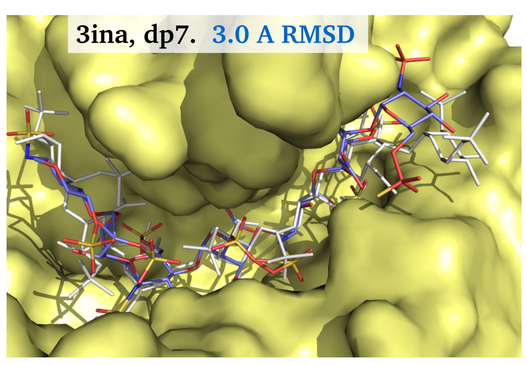


For each GAG we tested to assemble all sub-chains with dp5 to 7.

We could sample near-native (<5Å) dp5 chains for 12/13 complexes, and quasi-native (<3Å) for 6/13 complexes.

Those ratio diminished to 6/11 for **dp6, and** 2/11 for **dp7 chains**.

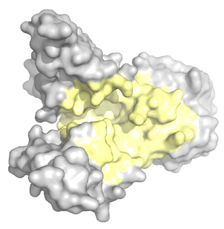




METHODS

Inputs

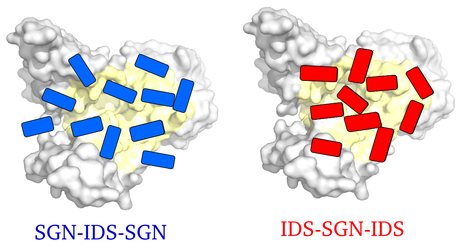
We chose 3mer as fragment length in order to conciliate binding specificity (enough contacts per frag.) and a low number of flexible bonds.
GAG having a periodicity in [A-B]n, we dock two types of trimers : A-B-A and B-A-B



The binding region can be predicted by fast rigid docking of GAG fragments and selection of the most contacting residues [5], or by electrostatic potential calculations [6], with good precision.

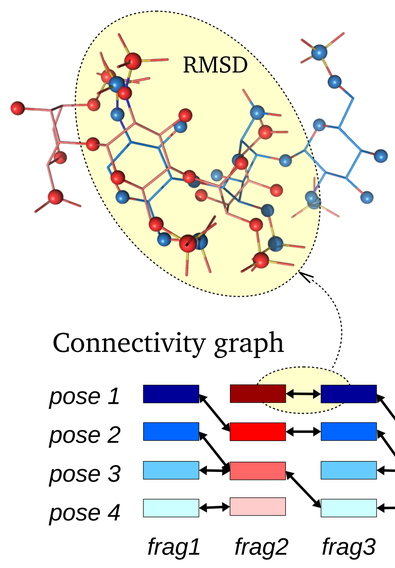
Here we consider the binding site as coarsely known, covering ~half of the surface.

Flexible Docking

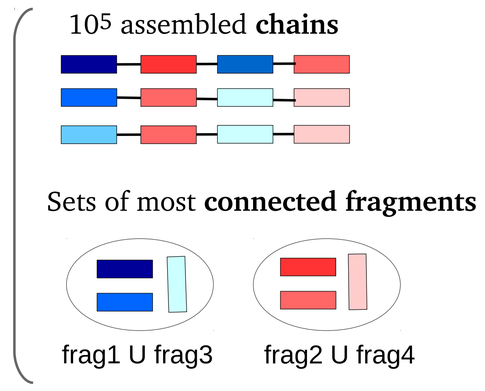


- **AutoDock3** [Morris *et al.* 1999]
- Charges : GLYCAM06 ff and literature (for SO4)
- All atoms representation, implicit solvent
- Grid centered on the COM of the bound ligand
- 1000 poses for each fragment type
- Rigid receptor (bound), fully flexible ligand

Combinatorial assembly

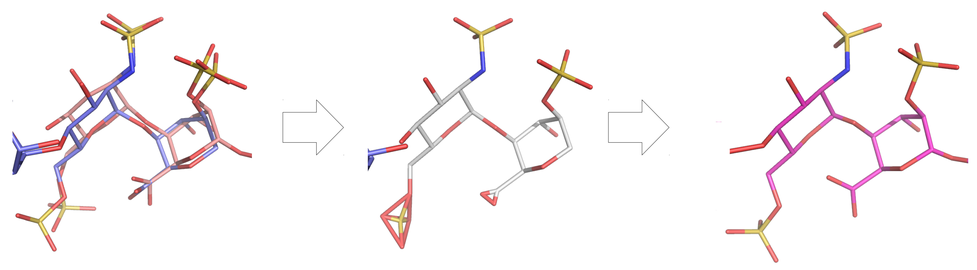


Pose compatibility depends on their overlapping RMSD. For time saving, only some atoms are considered (S/O/N). The cutoff is chosen so as to get at least 1000 chains for 5mers or 10.000 for 6/7mers respectively, up to a maximum of 3 Å.



Refinement

The assembled chains are clustered (0.5A) and the overlapping parts of the compatible poses are averaged. This results in incorrect monomer geometries, which is corrected by replacing each monomer by the closest one in a structural library, extracted from all initial docking poses.



CONCLUSION

We developed a new method to model protein-bound GAGs with high accuracy, on a coarsely known binding site on the protein. The approach proved effective to select correct fragment poses at the surface of the protein, and to model GAG up to a degree of polymerisation **dp7** with an accuracy of **3.0 Å RMSD**, with more than **10% correct models**.

Perspectives

- We considered the binding region as coarsly known. We will extend the method to cases were the **binding region could not be predicted** with sufficient confidence. We could either repeat the AD docking with a large number of grids, or use another docking engine for exhaustive sampling.
- For this second option, we will implement a **saccharide coarse-grained representation in the ATTRACT docking program**.
- We considered the protein as rigid. We intend to model also the **protein flexibility**, by representing parts of the surface in several conformations, that will have to match with neighbor conformations in the chain assembly.
- Finally, we will apply the method to the docking of **intrinsically disordered proteins (IDPs)**

[1] J. Kreuger *et al.* (2006) *J Cell Biol.* [2] D. Scharnweber *et al.* (2015) *JMSJ*
[3] C. de Beauchêne, S.J. de Vries, M. Zacharias (2016) *PloS.Comput.Biol.*
[4] C. de Beauchêne, S.J. de Vries, M. Zacharias (2016) *Nucleic Acids Research*
[5] N. Gandhi *et al* (2012) *Glycobiol* **[6] S. Samsonov and M.T. Pisabarro (2016) *Glycobiol***

The authors thanks the **Deutsche Forschungsgemeinschaft (DFG)** for funding part of this work, and Sjoerd J. De Vries for useful discussions